INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Well Balanced Designs for Data Placement

Jean-Claude Bermond — Alain Jean-Marie — Dorian Mazauric — Joseph Yu

N° 7725

August 2011

Thème COM

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--7725--FR+ENG

Well Balanced Designs for Data Placement

Jean-Claude Bermond* , Alain Jean-Marie* , Dorian Mazauric* , Joseph Yu*

Thème COM — Systèmes communicants
Projets Mascotte

Rapport de recherche n° 7725 — August 2011 — 14 pages

Abstract: The problem we consider in this article is motivated by data placement in particular data replication in video on demand systems. We are given a set V of n servers and b files (data, documents). Each file is replicated on exactly k servers. The problem is to determine the placement that minimizes the variance of the number of unavailable datas. To do that, we consider the problem of determining well balanced designs, a difficult problem because it contains the problem of the existence of Steiner systems.

Key-words: data placement, balanced designs, Steiner systems

This work was partially funded by Région PACA.

* MASCOTTE, INRIA, I3S, CNRS, Univ. Nice Sophia, Sophia Antipolis, France. `firstname.lastname@sophia.inria.fr`

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : +33 4 92 38 77 77 — Télécopie : +33 4 92 38 77 65

Configurations bien équilibrées pour le placement de données dans les réseaux pair-à-pair

Résumé : Nous considérons un problème motivé par la réplication des données dans un système de vidéo à la demande [BBJMR09b, JMRBB09] où chaque donnée est répliquée sur exactement k serveurs et où chaque serveur peut tomber en panne. Nous cherchons à déterminer un placement minimisant la variance du nombre de données indisponibles. Nous ramenons ce problème à un problème d'existence de configurations "équilibrées", problème difficile car il contient le problème de l'existence de système de Steiner.

Mots-clés : placement de données, configurations équilibres, systèmes de Steiner

1 Introduction

The problem we consider in this article is motivated by data placement in particular data replication in video on demand systems (see [BBJMR08, BBJMR09a, BBJMR09b, JMRBB09]). We use here the terminology of design and graph theory (so the notations are somewhat different from the papers mentioned above). We are given a set V of n servers and b files (data, documents). Each file is replicated (placed) on exactly k servers. The set of servers containing file i is therefore a subset of size k, which will be called a block and denoted B_i. A placement consists of giving a family $\mathcal{F}$ of blocks $B_i, 1 \leq i \leq b$.

A server is available (on-line) with some probability δ and so unavailable (offline, failed) with the probability $1-\delta$. The file i is said to be available if one of the servers containing it is available or equivalently the file is unavailable if all the servers containing it are unavailable. In [BBJMR09a, BBJMR09b, JMRBB09] the authors study the random variable Λ, the number of available files and they proved that the mean $E(\Lambda) = b(1-\delta^k)$; so the mean is independent of the placement. However thay proved that the variance of Λ depends on the placement and showed (see [JMRBB09]) that minimizing the variance corresponds to minimizing the polynomial $P(\mathcal{F}, x) = \sum_{j=0}^{k} v_j x^j$ where $x = \frac{1}{1-\delta}$ (so $x \geq 1$) and v_j denotes the number of ordered pairs of blocks intersecting in exactly j elements. So we can summarize our problem as follows:

Problem: Let n, k, b be given integers and x be a real number, $x \geq 1$; find a placement that is a family $\mathcal{F}$ of b blocks, each of size k, on a set of n elements, which minimizes the polynomial $P(\mathcal{F}, x) = \sum_{j=0}^{k} v_j x^j$, where v_j denotes the number of ordered pairs of blocks intersecting in exactly j elements. Such a placement will be called optimal for the value x.

In [JMRBB09] it is conjectured that for any n, k, b there exists a family $\mathcal{F}^*$ which is optimal for all the values of $x \geq 1$ (that is $P(\mathcal{F}^*, x) \leq P(\mathcal{F}, x)$ for any $\mathcal{F}$ and any $x \geq 1$).

Before stating our results let us give some examples. Let $n = 4, b = 4, k = 2$. We can consider different placements:

- Family $\mathcal{F}_1 : B_1 = B_2 = B_3 = B_4 = \{1, 2\}$; then $P(\mathcal{F}_1, x) = 12x^2$
- Family $\mathcal{F}_2 : B_1 = B_2 = \{1, 2\}, B_3 = B_4 = \{3, 4\}$; then $P(\mathcal{F}_2, x) = 4x^2 + 8$
- Family $\mathcal{F}_3 : B_1 = \{1, 2\}, B_2 = \{1, 3\}, B_3 = \{1, 4\}, B_4 = \{2, 3\}$; then $P(\mathcal{F}_3, x) = 10x+2$
- Family $\mathcal{F}_4 : B_1 = \{1, 2\}, B_2 = \{2, 3\}, B_3 = \{3, 4\}, B_4 = \{1, 4\}$; then $P(\mathcal{F}_4, x) = 8x+4$.

For any $x \geq 1$, $P(\mathcal{F}_4, x) \leq P(\mathcal{F}_i, x)$ and it can be proven that indeed $\mathcal{F}_4$ is an optimal family for any $x \geq 1$. Note that according to the values of x, $\mathcal{F}_2$ can be better (or worse) than $\mathcal{F}_3$. For $x \leq \frac{3}{2}$, $P(\mathcal{F}_2, x) \leq P(\mathcal{F}_3, x)$ (for example for $x = \frac{5}{4}$, $P(\mathcal{F}_2, \frac{5}{4}) = 14 + \frac{1}{4}$ and $P(\mathcal{F}_3, \frac{5}{4}) = 14 + \frac{1}{2}$). But for $x \geq \frac{3}{2}$, $P(\mathcal{F}_2, x) \geq P(\mathcal{F}_3, x)$ (for example for $x = 2$, $P(\mathcal{F}_2, 2) = 24$ and $P(\mathcal{F}_3, 2) = 22$).

Let now $n = 5, b = 3, k = 3$. We claim that the family $\mathcal{F}^*$ consisting of the three blocks $\{1, 2, 3\}, \{1, 2, 4\}, \{3, 4, 5\}$ is optimal for all $x \geq 1$. We have $P(\mathcal{F}^*, x) = 2x^2 + 4$. Let $\mathcal{F}$ be any other family with a polynomial $P(\mathcal{F}, x) = \alpha x^3 + \beta x^2 + \gamma x + \delta$. As $n = 5$, there can never be two disjoint blocks; so $\delta = 0$. Furthermore we always have $\alpha+\beta+\gamma = b(b-1) = 6$. So $P(\mathcal{F}, x) - P(\mathcal{F}^*, x) = (x-1)(\alpha x^2 + (\alpha + \beta - 2)x)$. If $\alpha \geq 2$ (that is at least one block repeated), then $P(\mathcal{F}, x) - P(\mathcal{F}^*, x) > 0$ for any $x > 1$. If $\alpha = 0$, among 3 blocks necessarily two of them have a pair in common and so $\beta \geq 2$ and $P(\mathcal{F}, x) - P(\mathcal{F}^*, x) \geq 0$ for any $x > 1$.

2 Our results

For a family $\mathcal{F}$ let $\lambda^{\mathcal{F}}_{x_1,\dots,x_j}$ (or shortly $\lambda_{x_1,\dots,x_j}$) denote the number of blocks of the family containing the j-element subset $\{x_1, \dots, x_j\}$. We first show that $P(\mathcal{F}, x) = \Sigma_{j=1}^{k} \Sigma_{x_1,\dots,x_j} \lambda^2_{x_1,\dots,x_j}(x-1)^j - bx^k + b^2$. A family $\mathcal{F}$ is j-balanced if the $\lambda_{x_1,\dots,x_j}$ are all equal or almost equal that is if for any two j-elemnt subsets $\{x_1, \dots, x_j\}$ and $\{y_1, \dots, y_j\}$ $|\lambda_{x_1,\dots,x_j} - \lambda_{y_1,\dots,y_j}| \leq 1$. Finally the family $\mathcal{F}$ is well balanced if it is j-balanced for $1 \leq j \leq k$. The form of the above polynomial enables us to prove that a well balanced family is also optimal and therefore the conjecture is proven for the values of b, for which there exists a well balanced family. The rest of the paper is devoted to construct well balanced families and so optimal ones. We consider first the case $k = 2$ where such families are easy to construct for any b. Then, we partly deal with the case $k = 3$ using results of design theory. Indeed the problem of constructing well balanced families contains as subproblem the existence of Steiner systems. Recall that a t-Steiner system satisfies(see [CM06]) the fact that, for any $1 \leq j \leq t$, $\lambda_{x_1,\dots,x_j}$ is a constant. As example a $(n, 3, 1)$ Steiner triple system is defined as a family of triples (blocks of size 3) such that every pair of elements belongs to exactly one block ($\lambda_{x_1,x_2} = 1$). So it is 2 balanced; it is well known that every vertex belongs to exactly $\frac{n-1}{2}$ blocks and therefore it is well balanced. Such a design exists if and only if $n \equiv 1$ or $3 \pmod 6$. In that case $b = \frac{n(n-1)}{6}$. That gives some sporadic values for which there exist well balanced families. We construct many other families; as example we show that such families exist for any b for the values of $n \equiv 3$ (mod6) for which there exist a large number of disjoint Kirkman triple systems (see []). We deal also with other congruences of n.

3 Computation of $P(\mathcal{F}, x)$

Recall that $\lambda_{x_1,\dots,x_j}$ denotes the number of blocks of the family containing the j-element subset $\{x_1, \dots, x_j\}$. By convention $\lambda_\emptyset = b$

Proposition 1. $P(\mathcal{F}, x) = \sum_{j=0}^{k} \sum_{x_1,\dots,x_j} \lambda_{x_1,\dots,x_j}(\lambda_{x_1,\dots,x_j} - 1)(x-1)^j$

Proof. $P(\mathcal{F}, x) = \sum_{h=0}^{k} v_h x^h$. Let us write $P(\mathcal{F}, x) = \sum_{j=0}^{k} \mu_j (x-1)^j$. Using $x^h = (x-1+1)^h = \sum_{j=0}^{h} \binom{h}{j}(x-1)^j$, we get $\mu_j = \sum_{h=j}^{k} \binom{h}{j} v_h$.
We claim that $\mu_j = \sum_{x_1,\dots,x_j} \lambda_{x_1,\dots,x_j}(\lambda_{x_1,\dots,x_j} - 1)$.

Indeed $\lambda_{x_1,\dots,x_j}(\lambda_{x_1,\dots,x_j}-1)$ counts the number of ordered pairs of blocks which contain $x_1,\dots,x_j$.This number is the sum of the ordered pairs of blocks which intersect in exactly the j elements $x_1,\dots,x_j$, plus those intersecting in exactly $j+1$ elements containing $x_1,\dots,x_j$, plus more generally those intersecting in exactly in h elements containing $x_1,\dots,x_j$, where, $j+2\le h\le k$. When we sum on all the possible j-elemnt subsets, we therefore get

- the number of ordered pairs of blocks intersecting in exactly j elements, that is v_j
- plus the number of ordered pairs of blocks intersecting in exactly $j+1$ elements, which are counted $\binom{j+1}{j}$ times. Indeed, if the intersection of two blocks is $\{x_1,\dots,x_{j+1}\}$ they are counted for all the j-element subsets included in $\{x_1,\dots,x_{j+1}\}$ which are in number $\binom{j+1}{j}$. Therefore we have $\binom{j+1}{j}v_{j+1}$ such ordered pairs of blocks.
- plus for a general $h, j+2\le h\le k$ we count $\binom{h}{j}v_h$ ordered pairs of blocks intersecting in exactly h elements; indeed if the intersection of two blocks is $\{x_1,\dots,x_h\}$ they are counted for all the j-element subsets included in $\{x_1,\dots,x_h\}$ which are in number $\binom{h}{j}$.

□

We will use intensively the following equality

$$\sum_{x_1,\dots,x_j}\lambda_{x_1,\dots,x_j}=b\binom{k}{j} \quad (1)$$

It follows from the fact that a given block B is counted once in all the $\lambda_{x_1,\dots,x_j}$ such that $\{x_1,\dots,x_j\}\subset B$ and we have $\binom{k}{j}$ such j-element subsets.

Theorem 1. $P(\mathcal{F},x)=\sum_{j=1}^{k}\sum_{x_1,\dots,x_j}\lambda^2_{x_1,\dots,x_j}(x-1)^j-bx^k+b^2$

Proof. Using equation 1, we get $\sum_{j=0}^{k}\sum_{x_1,\dots,x_j}\lambda_{x_1,\dots,x_j}(x-1)^j=\sum_{j=0}^{k}b\binom{k}{j}(x-1)^j=bx^k$. Replacing in the expression of $P(\mathcal{F},x)$ given in proposition 1 and using the fact that $\lambda^2_{\emptyset}=b^2$ we obtain the theorem. □

4 Well balanced families

A family $\mathcal{F}$ is j-balanced if all the $\lambda_{x_1,\dots,x_j}$ are equal or almost equal that is if for any two j-element subsets $\{x_1,\dots,x_j\}$ and $\{y_1,\dots,y_j\}$ $|\lambda_{x_1,\dots,x_j}-\lambda_{y_1,\dots,y_j}|\le 1$. Finally the family $\mathcal{F}$ is well balanced if it is j-balanced for $1\le j\le k$.

Proposition 2. *$\sum_{x_1,\dots,x_j}\lambda^2_{x_1,\dots,x_j}$ is minimized when $\mathcal{F}$ is j-balanced.*

Proof. As by equation 1, $\sum_{x_1,\ldots,x_j} \lambda_{x_1,\ldots,x_j}$ is a constant, $\sum_{x_1,\ldots,x_j} \lambda^2_{x_1,\ldots,x_j}$ is minimized when all the $\lambda_{x_1,\ldots,x_j}$ are equal to $b\frac{\binom{k}{j}}{\binom{n}{j}}$ if this value is an integer or equal either to $\lfloor b\frac{\binom{k}{j}}{\binom{n}{j}} \rfloor$ or $\lceil b\frac{\binom{k}{j}}{\binom{n}{j}} \rceil$. That is equivalent to say that $\mathcal{F}$ is j-balanced. □

So, we can state our main theorem

Theorem 2. *If $\mathcal{F}^*$ is well balanced, then $\mathcal{F}^*$ is optimal that is $P(\mathcal{F}^*, x) \leq P(\mathcal{F}, x)$ for any $\mathcal{F}$ and any $x \geq 1$.*

Proof. If $\mathcal{F}^*$ is well balanced, then all the coefficients of the polynomial as expressed in the Theorem 1 are minimized and so $\mathcal{F}^*$ is optimal. □

Note that for a j-balanced family, the coefficient of $(x-1)^j$ in the polynomial $P(\mathcal{F}, x)$ is easy to compute. Let $b\binom{k}{j} = q\binom{n}{j} + r$, then we have r values of the $\lambda_{x_1,\ldots,x_j}$ equal to $q+1$ and $\binom{n}{j} - r$ equal to q. So, $\sum_{x_1,\ldots,x_j} \lambda^2_{x_1,\ldots,x_j} = \binom{n}{j}q^2 + 2qr + r$.

Complete well balanced family:

If $b = \binom{n}{k}$ the complete family consisting of all the possible k-subsets is well balanced, with the values of the $\lambda_{x_1,\ldots,x_j}$ being all equal to $\lambda_j = \binom{n-j}{k-j}$. By taking h copies we get also a well balanced family for $b = h\binom{n}{k}$.

Proposition 3. *Let n and k be given and let $b' = h\binom{n}{k} + b$ with $b \leq \binom{n}{k}$. Then, there exists a well balanced family $\mathcal{F}'$ for b' if and only if there exists a well balanced family $\mathcal{F}$ for b*

Proof. If we have a well balanced family $\mathcal{F}$ for some $b \leq \binom{n}{k}$ we can construct a well balanced family $\mathcal{F}'$ for $b' = h\binom{n}{k} + b$ by adding h complete families to $\mathcal{F}$. Conversely if we have a well balanced family $\mathcal{F}'$ for $b' = h\binom{n}{k} + b$, each k-element subset is repeated h or $h+1$ times and so by deleting h copies of each block, we can deduce a well balanced family for b. □

The next proposition generalizes this idea to optimal families.

Proposition 4. *Let n and k be given and let $b' = h\binom{n}{k} + b$ with $b \leq \binom{n}{k}$. If there exists an optimal family for b', then there exists an optimal family for b and furthermore the optimal family for b' consists of the optimal family for b plus h complete families.*

Proof. Suppose there exists an optimal family $\mathcal{F}'$ for b'. This family is necessarily k-balanced. Indeed suppose it is not the case and let $\mathcal{G}'$ be a k-balanced family (such a family can be easily constructed by taking among the $\binom{n}{k}$ subsets of size k, b of them repeated $h+1$ times and the other $\binom{n}{k} - b$ repeated h times). But, the coefficient of x^k in $P(\mathcal{G}', x)$ will be strictly less than that of $P(\mathcal{F}', x)$ and so for x large enough $P(\mathcal{G}', x) < P(\mathcal{F}', x)$ contradicting the optimality of $\mathcal{F}'$. So each k-element subset appears exactly h or $h+1$ times. Deleting h

copies of each block we get a family $\mathcal{F}$ with $b = b' - \binom{n}{k}$ blocks (none of them being repeated). Note that if $\lambda_{x_1,\ldots,x_j}$ (resp $\lambda'_{x_1,\ldots,x_j}$) denotes the number of blocks of the family $\mathcal{F}$ (resp $\mathcal{F}'$) containing $\{x_1,\ldots,x_j\}$ we have : $\lambda'_{x_1,\ldots,x_j} = \lambda_{x_1,\ldots,x_j} + h\binom{n-j}{k-j}$. Consider another family $\mathcal{G}$ on b blocks and let $\mathcal{G}'$ be the family on b' blocks obtained by adding h complete families to $\mathcal{G}$. Let $\mu_{x_1,\ldots,x_j}$ (resp $\mu'_{x_1,\ldots,x_j}$) denote the number of blocks of the family $\mathcal{G}$ (resp $\mathcal{G}'$) containing $\{x_1,\ldots,x_j\}$. Then we have : $\mu'_{x_1,\ldots,x_j} = \mu_{x_1,\ldots,x_j} + h\binom{n-j}{k-j}$. So, as by equation 1, $\sum_{x_1,\ldots,x_j}\lambda_{x_1,\ldots,x_j} = \sum_{x_1,\ldots,x_j}\mu_{x_1,\ldots,x_j}$ and $\sum_{x_1,\ldots,x_j}\lambda'_{x_1,\ldots,x_j} = \sum_{x_1,\ldots,x_j}\mu'_{x_1,\ldots,x_j}$, then $\sum_{x_1,\ldots,x_j}\lambda^2_{x_1,\ldots,x_j} - \sum_{x_1,\ldots,x_j}\mu^2_{x_1,\ldots,x_j} = \sum_{x_1,\ldots,x_j}\lambda'^2_{x_1,\ldots,x_j} - \sum_{x_1,\ldots,x_j}\mu'^2_{x_1,\ldots,x_j}$ and thus $P(\mathcal{G}',x) - P(\mathcal{F}',x) = P(\mathcal{G},x) - P(\mathcal{F},x)$. Therefore if $\mathcal{F}$ is not optimal there exists a family $\mathcal{G}$ and a value x for which $P(\mathcal{G},x) < P(\mathcal{F},x)$ and for this value of x we have $P(\mathcal{G}',x) < P(\mathcal{F}',x)$ and $\mathcal{F}'$ will not have been optimal, a contradiction. □

Note that we conjecture that the converse is true: that is starting from an optimal family $\mathcal{F}$ for some $b \leq \binom{n}{k}$, the family $\mathcal{F}'$ obtained by adding h complete families is also optimal. That is verified, if the conjecture of [JMRBB09]on the existence of an optimal family for any n, b, k is true as in that case any optimal family is k-balanced.

In what follows we will restrict ourselves to the case $b \leq \binom{n}{k}$. In fact the following proposition shows that we can consider only the values of $b \leq \frac{1}{2}\binom{n}{k}$.

Proposition 5. *Let n and k be given, an optimal family $\bar{\mathcal{F}}$ for $\bar{b} = \binom{n}{k} - b$ is obtained from an optimal family $\mathcal{F}$ for $b \leq \binom{n}{k}$ by taking as blocks the k-subsets which are not blocks of $\mathcal{F}$.*

Proof. Let $\mathcal{F}$ be an optimal family with b blocks and let $\bar{\mathcal{F}}$ be the family obtained from $\mathcal{F}$ by taking as blocks the k-subsets which are not blocks of $\mathcal{F}$. $\bar{\mathcal{F}}$ has $\bar{b} = \binom{n}{k} - b$ blocks. Furthermore, if $\bar{\lambda}_{x_1,\ldots,x_j}$ denotes the number of blocks of the family $\bar{\mathcal{F}}$ containing $\{x_1,\ldots,x_j\}$, we have $\bar{\lambda}_{x_1,\ldots,x_j} = \binom{n-j}{k-j} - \lambda_{x_1,\ldots,x_j}$. Consider another family $\bar{\mathcal{G}}$ with $\bar{b}$ blocks and let $\mathcal{G}$ be the complementary family obtained from $\bar{\mathcal{G}}$ by taking as blocks the k-subsets which are not blocks of $\bar{\mathcal{G}}$; $\mathcal{G}$ has b blocks. We also have: $\bar{\mu}_{x_1,\ldots,x_j} = \binom{n-j}{k-j} - \mu_{x_1,\ldots,x_j}$ and so we get $P(\bar{\mathcal{G}},x) - P(\bar{\mathcal{F}},x) = P(\mathcal{G},x) - P(\mathcal{F},x)$. Therefore if $\mathcal{F}$ is an optimal family, then $\bar{\mathcal{F}}$ is also an optimal family.

□

Well balanced families and Steiner systems:

Recall that a t-Steiner system (or (n,k,λ) t-design) is a family of blocks such that each t-element subset appears in exactly λ blocks (see [CM06, CJ06]). In that case it is well known that also, for $1 \leq j \leq t$ each j-element subset appears in exactly λ_j blocks, where $\lambda_j = \frac{\lambda}{t+1-j}\binom{n-j}{t-j}$. So a t-design is j-balanced for all j, $1 \leq j \leq t$. In particular, if $t = k-1$ and the blocks are repeated the same or almost the same number of times, then a k-Steiner family is also well balanced. As an example, a Steiner Triple System (STS) consists of a family of triples, such that each pair of elements appears in exactly one triple. In that case each element appears in $\frac{n-1}{2}$ triples and no triple is repeated. Therefore, a STS is a well balanced family. It is well known that a STS exists if and only if $n \equiv 1$ or 3 mod (6) and then $b = \frac{n(n-1)}{6}$.

5 Case k = 2

Theorem 3. *Let $k = 2$. Then for any n and b there exists a well balanced family.*

Proof. We have only to consider the case $b \leq \binom{n}{2}$. In the case $k = 2$ the blocks are pairs of elements and so the problem consists of designing a simple graph with b edges and almost regular (the degree of a vertex x being $d(x) = \lfloor \frac{2b}{n} \rfloor$ or $\lceil \frac{2b}{n} \rceil$. We distinguish two cases:

- Case n even: let $b = q\frac{n}{2} + r$ for $0 \leq r < \frac{n}{2}$. It is well known that, for n even, the edges of the complete graph K_n can be decomposed into $n-1$ perfect matchings (set of $\frac{n}{2}$ disjoint edges covering the vertices). In that case the family consisting of q perfect matchings plus r edges of the $(q+1)$th perfect matching form the required family with $b = q\frac{n}{2} + r$ edges, none of them repeated and with the degree of a vertex equal to q or $q + 1$.

- Case n odd: let $b = qn + r$ for $0 \leq r < n$. It is also well known that for n odd, the edges of complete graph K_n can be decomposed into $\frac{n-1}{2}$ hamiltonian cycles (cycles containing each vertex exactly once). In that case consider the family consisting of q hamiltonian cycles plus the following r edges of the $(q + 1)$th hamiltonian cycle: if the cycle is $x_0, x_1, \ldots, x_i, \ldots, x_{n-1}$ we take the r edges $\{x_{2j}, x_{2j+1}\}$ for $0 \leq j \leq r - 1$ (indices being taken modulo n). Then it consists of $b = qn + r$ edges none of them being repeated; furthermore the degree of a vertex is $2q$ or $2q+1$ if $r \leq \frac{n-1}{2}$ and $2q+1$ or $2q + 2$ otherwise and so in both cases $d(x) = \lfloor \frac{2b}{n} \rfloor$ or $\lceil \frac{2b}{n} \rceil$.

□

Algorithm to construct a well balanced family starting from any family:
In some cases (files or servers appearing or disappearing), it might be helpful to design an algorithm, which starting from some family constructs an optimal well balanced family. That is in general a difficult problem; but for $k = 2$, we can easily design such a procedure. Let n and b be given and $k = 2$ and consider any family $\mathcal{F}$; we will transform it into a well balanced family with the same parameters. First let us construct a 2-balanced family. Suppose, $\mathcal{F}$ is not 2 balanced; so there exist two edges (blocks) $\{x, y\}$ and $\{z, t\}$ with $\lambda_{xy} \geq \lambda_{zt} + 2$. Then, delete from $\mathcal{F}$ one edge $\{x, y\}$ and add one edge $\{z, t\}$. Repeating this procedure we end after a finite number of steps with a family such that for any pair of edges $\{x, y\}$ and $\{z, t\}$ $\lambda_{xy} \leq \lambda_{zt} + 1$, that is a 2-balanced family.
Now let us show how to construct a well balanced family from a 2- balanced one. Let $\mathcal{F}$ be a 2-balanced family with $\lambda_{xy} = \lambda$ or $\lambda - 1$; suppose it is not 1-balanced; then there exist two

vertices x and z with $d(x) \geq d(z) + 2$. So there exists a vertex $y \neq x, z$ with $\lambda_{xy} \geq \lambda_{zy} + 1$; otherwise $d(x) = \sum_{y \neq x,z} \lambda_{xy} + \lambda_{xz} \leq \sum_{y \neq x,z} \lambda_{zy} + \lambda_{xz} = d(z)$ a contradiction. So, $\lambda_{xy} = \lambda$ and $\lambda_{zy} = \lambda - 1$. Deleting from $\mathcal{F}$ one edge $\{x, y\}$ and adding one edge $\{z, y\}$, we still get a 2-balanced family $\mathcal{F}'$ ($\lambda'_{xy} = \lambda - 1$ and $\lambda'_{zy} = \lambda$); but we have reduced the gap between the degrees of x and z (as $d'(x) = d(x) - 1$ and $d'(z) = d(z) - 1$ the other degrees being unchanged). Repeating this procedure we end after a finite number of steps with a 1-balanced and 2-balanced, so well balanced family.

6 Case k = 3

The cases $k > 2$, are much more complicated. Already for $k = 3$, there are values of n and b for which there do not exist well balanced families. For $n = 4$ and $b = 2$, if there exists a 2-balanced family, then $\lambda_{xy} = 1$, but that is impossible as $n - 1$ =3 and there cannot exist a partition of the edges of K_4 into triples. The argument is generalized in the following proposition :

Proposition 6. *Let $k = 3$, n be even and λ be odd. If $\lambda\frac{n(n-1)}{2} - \frac{n}{2} < 3b < \lambda\frac{n(n-1)}{2} + \frac{n}{2}$ then there does not exist a 2 balanced family.*

Proof. Note that the number of possible pairs is $\frac{n(n-1)}{2}$. We distinguish 3 cases

- $\lambda\frac{n(n-1)}{2} = 3b$. In that case a 2-balanced family will verify $\lambda_{xy} = \lambda$ for all pairs $\{x, y\}$ and then we should have $\lambda_x = \lambda\frac{n-1}{2}$ impossible as λ is odd and n is even.
- $\lambda\frac{n(n-1)}{2} - \frac{n}{2} < 3b$. In that case we cannot have all the $\lambda_{xy} \geq \lambda$. So one of the $\lambda_{xy} \leq \lambda - 1$ and if the family is 2-balanced all the $\lambda_{xy} \leq \lambda$. But, then $\lambda_x \leq \lambda\frac{n-1}{2}$ and as $\lambda\frac{n-1}{2}$ is odd, $\lambda_x \leq \lambda\frac{n-1}{2} - \frac{1}{2}$. Therefore, using Equation 1, $3b = \sum_x \lambda_x \leq \lambda\frac{n(n-1)}{2} - \frac{n}{2}$ a contradiction.
- $3b < \lambda\frac{n(n-1)}{2} + \frac{n}{2}$. In that case we cannot have all the $\lambda_{xy} \leq \lambda$. So one of the $\lambda_{xy} \geq \lambda + 1$ and if the family is 2-balanced all the $\lambda_{xy} \geq \lambda$. But, then $\lambda_x \geq \lambda\frac{n-1}{2}$ and as $\lambda\frac{n-1}{2}$ is odd, $\lambda_x \geq \lambda\frac{n-1}{2} + \frac{1}{2}$. Therefore, using Equation 1, $3b = \sum_x \lambda_x \geq \lambda\frac{n(n-1)}{2} - \frac{n}{2}$ a contradiction.

□

Examples of application : By Proposition 6, there do not exist well balanced families for$k = 3$ and $n = 6; b = 5$; $n = 8; b = 9, 10, 27, 28, 29$; $n = 10; b = 14, 15, 16, 44, 45, 46$

On a positive side we can use all the results obtained in design theory in particular on Steiner Triple Systems (see the handbook [CJ06] for details) to construct some well balanced families. Recall that a $(n, 3, 1)$ Steiner Triple System (STS(n) shortly) is defined as a family of triples (blocks of size 3) such that every pair of elements belongs to exactly one block ($\lambda_{x_1,x_2} = 1$). So it is 2 balanced; it is well known that every vertex belongs to exactly $\frac{n-1}{2}$

blocks and therefore it is well balanced. Such a design exists if and only if $n \equiv 1$ or 3 (mod 6). In that case $b = \frac{n(n-1)}{6}$. That gives some sporadic values for which there exist well balanced families. We can get more values of b by considering more than one STS(n); but we have to insure that the family is 3-balanced (that is no block are repeated). Fortunately the answer is obtained thanks to Theorem 4. Two STS(n) are said to be disjoint if they have no triple in common. A set of $n-2$ disjoint STS(n) is called a *large set of disjoint STS(v)* and briefly denoted by LSTS(n). A LSTS(n)can be viewed as a partition of the complete family of $\binom{n}{3}$ triples into STS(n). In 1850, Cayley showed that there are only two disjoint STS(7) and so there is no LSTS(7). The same year Kirkman showed that there exists an LSTS(9). Then, due to the efforts of many authors the following theorem completely settles the existence of LSTS(n).

Theorem 4. *([Lu83, Lu84, Tei91] (see [Ji05] for a simple proof) For $n \equiv 1$ or 3 (mod 6) with $n > 7$ there exits an LSTS(n).*

Proposition 7. *Let $k = 3$, and $n \equiv 1$ or 3 $n > 7$, then there exists a well balanced family for any b multiple of $\frac{n(n-1)}{6}$.*

Proof. Let $b = h\frac{n(n-1)}{6}$; $b \leq \binom{n}{3}$. Then, the family consisting of h dijoint STS(n) is well balanced (with $\lambda_{xy} = h$ and $\lambda_x = h\frac{n-1}{2}$). For $b \geq \binom{n}{3}$ the result follows by using Proposition 3 □

We will see after that the existence of two disjoint STS(7) suffices to construct a well balanced family for $n = 7$ and any b.

When $n = 6t + 3$) there exist STS which have a strong property. The triples of the STS can be themselves be partitionned into $3t+1$ classes, called parallel classes, where each class consists of $2t + 1$ blocks forming a partition of the n elements. Such an STS is called resolvable or a Kirkman Triple System (briefly KTS(n)).It is well known that a KTS(n) exists for any $n \equiv 3$ (mod6) ([RCW71]. Two KTS(n) are said to be disjoint if they have no triple in common. A set of $n-2$ disjoint KTS(n) is called a *large set of disjoint KTS(n)* and briefly denoted by LKTS(n). Kirkman showed that an LKTS(9) exists in 1850 and Denniston found an LKTS(15) in 1974. Since that many peoplke have done some research on their existence. The more recent paper is that of [ZC10] where the reader can find other references. The results are summarized in the following theorem :

Theorem 5. *[ZC10] There exists an LKTS$(3^a 5^b r \prod_{i=1}^{s}(2.13^{n_i} + 1 \prod_{j=1}^{t}(2.7^{m_j} + 1))$ for any integer $r \in \{7, 13\}, n_i, m_j \geq 1 (1 \leq i \leq s, 1 \leq j \leq t), a \geq 1, b, s, t \geq 0$ and further $a + s + t \geq 2$ if $b \geq 1$*

Proposition 8. *Let $k = 3$, and $n = 6t + 3$. If there exists an LKTS(n) then there exists a well balanced family for any b.*

Proof. By Proposition 3 we can suppose $b \leq \binom{n}{3}$. Let $b = q(2t+1)(3t+1)+r(2t+1)+s$ with $0 \leq q < 6t+1; 0 \leq r < 3t+1; 0 \leq s < 2t+1$. Let $K_i, 1 \leq i \leq 6t+1$ be the $6t+1$ KTS(n) of an LKTS(n). Then a well balanced family consists of q disjoint KTS plus r parallel classes of the $(q+1)$th KTS K_{q+1} and s triples of the $(r+1)$th parallel class of K_{q+1}. Indeed by definition of an LKTS, all the triples are disjoint and so $\lambda x, y, z = 0$ or 1. In eack KTS a pair of elements appears exactly once; so $\lambda x, y = q$ or $q+1$ (exactly q if $r = 0, s = 0$). In each parallel class, each vertex appears exactly once ; so $\lambda x = (3t+1)q + r$ or $(3t+1)q + r + 1$ (exactly $(3t+1)q + r$ if $s = 0$). □

Note that, according to the proof above and the fact that by Proposition 5 we can suppose $b \leq \frac{1}{2}\binom{n}{k}$, we do not need to have a strong structure such as an LKTS, but only $3t+1$ disjoint STS with one of them being a KTS (the one used as the $(q+1)$th KTS. We conjecture such structure always exists for $n = 6t+3$.

Conjecture 1. *For $n = 6t+3$ there exist $3t+1$ disjoint STS(n) one of them being a KTS(n)*

Concerning well balanced families we conjecture that:

Conjecture 2. *Let $k = 3$ and $n \equiv 1$ or 3 (mod6). Then there exists a well balanced family for any b*

Conjecture 2 will follow from Conjecture 1 for $n \equiv 3 \pmod 6$. For $n \equiv 1 \pmod 6$ we need to have some extra property. We will see that the conjecture is true for $n = 7$.

From Proposition 7 and Proposition 8 we can construct well balanced families for $n = 6t$ or $6t+2$ by deleting an element or for $n = 6t+2$ or $6t+4$ by adding one element; but we have to do it carefully.

Proposition 9. *Let $k = 3$ and $n = 6t$ (resp $n = 6t+2$). There exist a well balanced family for $b = ht(6t-2)$ (resp $b = ht(6t+2)$.*

Proof. Take a set of h disjoint STS(n+1) (as $n+1 \equiv 1$ or 3 (mod6)) and delete all the $h\frac{n}{2}$ blocks containing the element $n+1$. □

Proposition 10. *Let $k = 3$ and $n = 6t+4$. There exist a well balanced family for $b = 2p(3t+2)(2t+1)$*

Proof. We will use the following construction (called Construction A). Consider a parallel class of a KTS($6t+3$) and a new element $\alpha (= 6t+4)$ and replaced each of the $2t+1$ triples $\{x_j, y_j, z_j\}$ ($1 \leq j \leq 2t+1$ with the 3 triples $\{x_j, y_j, \alpha\}$, $\{x_j, z_j, \alpha\}$ and $\{y_j, z_j, \alpha\}$. Now start with $q = 2p$ disjoint KTS on a set X of $6t+3$ elements and do the construction A for p classes of one KTS. As the classes are taken in the same KTS, α appears in $3p(2t+1)$ disjoint triples, where each vertex x appears exactly $2p$ times; so $\lambda_{\alpha x} = 2p$ and $\lambda_\alpha = 3p(2t+1)$. Doing so we have not changed the values of $\lambda_{xy} = 2p$ as we took $2p$ disjoint KTS; but λ_x has increased by p and it value is now $2p(3t+1) + p = 3p(2t+1)$. Therefore the family constructed is well balanced. □

Small cases

n=5

For $n = 5$, we have well balanced families for $b = 1$ (one block) and $b = 2$ (two blocks $\{1, 2, 3\}$ and $\{1, 4, 5\}$), but not for $b = 3$ as we have seen in the example of the introduction. However there exists an optimal solution $\{1, 2, 3\}, \{1, 2, 4\}, \{3, 4, 5\}$ 1-balanced but not 2-balanced ($\lambda_{12} = 2$ but $\lambda_{15} = \lambda_{25} = 0$). One can show also there is no well balanced solution for $b = 4$; an optimal one consists of the blocks $\{1, 2, 3\}, \{1, 2, 4\}, \{1, 3, 5\}, \{3, 4, 5\}$. For $b = 5$ there exists a well balanced solution with $\lambda_x = 3$ and $\lambda_{xy} = 1$ or 2 and consisting of the 5 blocks $\{1, 2, 3\}, \{1, 2, 4\}, \{1, 3, 5\}, \{3, 4, 5\}, \{2, 4, 5\}$.

n=6

For $n = 6$, $\binom{6}{3} = 20$ and by Proposition 5 we have to consider only the values of $b \leq 10$.

For $b = 5$ (and so $b = 15$) there does not exist Proposition 6 a well balanced family. An optimal solution $\mathcal{F}^*$ consists of the 5 blocks:
$\{1, 2, 3\}, \{1, 2, 4\}, \{1, 5, 6\}, \{2, 5, 6\}, \{3, 4, 5\}$ ($\lambda_x = 2$ or 3 and $\lambda_{12} = \lambda_{56} = 2$ but $\lambda_{36} = \lambda_{46} = 0$) with $4x^2 + 16x$ as associated polynomial. The proof is obtained by looking at various cases. Let a general solution be of the form $P(\mathcal{F}, x) = \alpha x^3 + \beta x^2 + \gamma x + \delta$. Furthermore, we always have $\alpha + \beta + \gamma = b(b - 1) = 20$. So $P(\mathcal{F}, x) - P(\mathcal{F}^*, x) = (x - 1)(\alpha(x - 1)^2 + (3\alpha + \beta - 4)(x - 1) + 2\alpha + \beta - 4 - \delta)$. If $\alpha \geq 2$ (that is at least one block repeated), one can show by looking at the number of repeated blocks that $P(\mathcal{F}, x) - P(\mathcal{F}^*, x) > 0$ for any $x > 1$. For example suppose one block is repeated 3 times say $\{1, 2, 3\}$; then either we have twice the block $\{4, 5, 6\}$ and then $\alpha = 8$ and $\delta = 12$ otherwise $\alpha = 6$ and $\delta \leq 6$.

In what follows, let $\alpha = 0$, that is no repeated block. Then $P(\mathcal{F}, x) - P(\mathcal{F}^*, x) = (\beta - 4)(x - 1) + \beta - 4 - \delta)$. Note that $\delta \leq 4$, as there can be at most two pairs of disjoint blocks (as $b = 5$); furthermore, when we have two disjoint blocks any other block intersect one block in 2 elements and the other in one element. We distinguish 3 cases :

- $\delta = 4$. Wlog let the pairs of disjoint blocks be $B_1 = \{1, 2, 3\}, B_2 = \{4, 5, 6\}, B_3 = \{1, 2, 4\}, B_4 = \{3, 5, 6\}$. Then B_1 and B_3 intersect in a pair and also B_2 and B_4; the last block B_5 intersect one of B_1, B_2 and one of B_3, B_4 in a pair; so $\beta = 8$ and $P(\mathcal{F}, x) - P(\mathcal{F}^*, x) \geq 0$ for $x \geq 1$.
- $\delta = 2$. Let the two disjoint blocks be B_1 and B_2. Then any other block intersect one of them in a pair and so $\beta \geq 6$ and $P(\mathcal{F}, x) - P(\mathcal{F}^*, x) \geq 0$ for $x \geq 1$.
- $\delta = 0$. First, $\beta = 0$ is impossible as if $B_1 = \{1, 2, 3\}$, we have at most 3 pairs available among $\{4, 5, 6\}$. If $\beta = 2$; let the two blocks intersecting in a pair be $B_1 = \{1, 2, 3\}, B_2 = \{1, 2, 4\}$; the pair $\{5, 6\}$ can be in at most one triple, but the two other blocks need to share either one pair with B_1 or B_2 or the pair $\{3, 4\}$. So necessarily $\beta \geq 4$ and $P(\mathcal{F}, x) - P(\mathcal{F}^*, x) \geq 0$ for $x \geq 1$.

For the other values of b, we can construct well balanced families as follows. Let $B_1 = \{1, 2, 3\}, B_2 = \{4, 5, 6\}$; $C_1 = \{1, 2, 4\}, C_2 = \{1, 3, 5\}, C_3 = \{2, 3, 6\}$; $D_1 = \{1, 4, 5\}, D_2 = \{2, 4, 6\}, D_3 = \{3, 5, 6\}$ and $C'_1 = \{1, 2, 5\}, C'_2 = \{1, 3, 6\}, C'_3 = \{2, 3, 4\}$. Note that the C_i and C'_i (resp D_i) intersect B_1 (resp. B_2) in three different pairs and B_2 (resp B_1) in 3 different elements. Solutions are obtained by taking:

- for $b = 1$ B_1;
- for $b = 2$, B_1, B_2;
- for $b = 3$, C_1, C_2, C_3;
- for $b = 4$, C_1, C_2, C_3, B_2;
- for $b = 6$, $C_1, C_2, C_3, D_1, D_2, D_3$;
- for $b = 7$, $C_1, C_2, C_3, D_1, D_2, D_3, B_1$;
- for $b = 8$, $C_1, C_2, C_3, D_1, D_2, D_3, B_1, B_2$;
- for $b = 9$, $C_1, C_2, C_3, D_1, D_2, D_3, C'_1, C'_2, C'_3$;
- for $b = 10$, $C_1, C_2, C_3, D_1, D_2, D_3, C'_1, C'_2, C'_3, B_1$.

n=7

For $n = 7$, $\binom{7}{3} = 35$; by Proposition 3 and Proposition 5 we have to consider only the values of $b \leq 17$.

Proposition 11. *For $k = 3$ and $n = 7$, there exists a balanced family for any b.*

Proof. Kirkman proved that there exist two disjoint STS(7). The first one consists of the 7 blocks $C_i = \{i+1, i+2, i+4\}$, for $0 \leq 6$ and the second one of the 7 blocks $D_i = \{i+1, i+3, i+4\}$, for $0 \leq 6$ (indices modulo 7). Let $B_1 = \{1,2,3\}, B_2 = \{4,5,6\}, B_3 = \{1,4,7\}, B_4 = \{2,5,7\}, B_5 = \{3,6,7\}$. For $b = j, 1 \leq j \leq 5$ take the blocks $B_i, 1 \leq i \leq j$. For $b = 7$ take the first STS(7) (that is all the C_i). For $b = 6$ delete one block from the STS(7). For $b = 7 + j, 1 \leq j \leq 5$ add to the STS(7) the blocks $B_i, 1 \leq i \leq j$. For $b = 14$ take the two disjoint STS(7) (that is all the C_i and D_i). For $b = 13$ delete one block from one STS(7). For $b = 14 + j, 1 \leq j \leq 5$ add to the two disjoint STS(7) the blocks $B_i, 1 \leq i \leq j$. □

References

[BBJMR08] A-E. Baert, V. Boudet, A. Jean-Marie, and X. Roche. Minimization of download times in a distributed vod system. In *ICPP08: The international conference on parallel processing*, pages 173–180, Los Alamitos, CA, USA, 2008. IEEE.

[BBJMR09a] A-E. Baert, V. Boudet, A. Jean-Marie, and X. Roche. Minimization of download time variance in a distributed vod system. *Scalable Computing Practice and experience*, 10(1):75–86, 2009.

[BBJMR09b] A-E. Baert, V. Boudet, A. Jean-Marie, and X. Roche. Performance analysis of dat replication in grid delivery networks. In *Int. Conf. on Complex Intelligent and Software Intensive Systems*, pages 369–374, 2009.

[CJ06] Colbourn C.J. and Dinitz J.H., editors. *The CRC Handbook of Combinatorial Designs (2nd edition)*, volume 42. CRC Press, 2006.

[CM06] C. J. Colbourn and R. Mathon. *The CRC Handbook of Combinatorial Designs (2nd edition)*, volume 42 of *Discrete Mathematics and Its Applications*, chapter *Steiner systems, pages 102–110. CRC Press, C.J. Colbourn and J.H. Dinitz edition, 2006.*

[Ji05] L. Ji. A new existence proof for large sets of disjoint steiner triple systems. J. Combinatorial theory A*, 112:308–327, 2005.*

[JMRBB09] A. Jean-Marie, X. Roche, A-E. Baert, and V. Boudet. Combinatorial designs and availability. Technical Report RR 7119, INRIA, 2009.

[Lu83] J.X. Lu. On large sets of disjoint steiner triple systems i,ii, iii. J. Combinatorial theory A*, 34:140–146,147–155, and 156–182, 1983.*

[Lu84] J.X. Lu. On large sets of disjoint steiner triple systems iv,v, vi. J. Combinatorial theory A*, 37:136–163,164–188, and 189–192, 1984.*

[RCW71] D.K. Ray-Chauduri and R.M. Wilson. Solution of kirkman's schoolgirl problem. Proc. symp. pure Math*, 19:187–204, 1971.*

[Tei91] L. Teirlinck. A completion of lu's determination of the spectrum of large sets of disjoint steiner triple systems. J. Combinatorial theory A*, 57:302–305, 1991.*

[ZC10] J. Zhou and Y. Chang. new results on large sets of kirkman triple systems. Des Codes Cryptogr*, 55:1–7, 2010.*

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399